

Kearsley, Lynch, and Wizer (1995) reviewed the literature that examines the use of computer conferencing in higher education and found that, in comparison with traditional classes, student satisfaction with online courses is higher, measures of student achievement are the same or better, and there is usually more discussion among students and instructors in a course.

Schutte (1997) reported on an experiment carried out during a fall 1996 social statistics course at California State University, Northridge, in which students were randomly divided into two groups—one taught in a traditional classroom and the other taught virtually on the World Wide Web. Text, lectures, and exams were standardized between the two groups. Schutte found that, contrary to expectations, the virtual class scored an average of 20 percent higher than the traditional class on both examinations.

At the same time, distance education raises issues concerning broader effects on the university. Although online education may expand the pool of people who have access to education, it also may take students away from traditional education, and some scholars express concern that it will undermine the traditional college experience. Some people question whether the quality of distance education can match that of face-to-face instruction. Moreover, creating the kind of intellectual or social community that characterize colleges may be much harder through distance learning.

Distance education also brings universities into competition with each other in a new way. Because distance education courses are available to anyone, anywhere, they allow universities to compete for students in other geographic areas. Top-tier universities such as Stanford and Duke are beginning to market Internet-based master's degrees to national audiences. New distance education-based universities—such as Jones International University (<<<http://www.jonesinternational.edu>>>), the first online-only university to gain accreditation; the University of Phoenix online (<<<http://online.uophx.edu>>>); and the Western Governors University (<<<http://www.wgu.edu>>>)—are also marketing courses that compete with universities and community colleges that have in the past been providers of continuing education services in their region. Others see opportunities to market American university degrees to large potential student populations abroad. The reverse is also happening. The United Kingdom's Open University, which has established a good reputation as a provider of distance education in the U.K. since 1971, has started an operation in the United States (Blumenstyk 1999a).

In addition, distance education is creating new markets for companies selling course materials and software to assist in online courses (Blumenstyk 1999b). Publishers such as McGraw-Hill and software companies such as Microsoft and Oracle have developed and are marketing online courses (Morris 1999).

Some people regard distance education technologies as providing new tools to professors. Others foresee mass production education, in which packaged multimedia courses will reduce the importance of professors (Noble 1998). As one indicator of concern, more than 850 faculty members at the

University of Washington signed a letter to Governor Gary Locke protesting the state's plans for investing in information technology (Monaghan 1998). The expanding and potentially lucrative new market for online course materials has also raised the issue of whether professors or the university should own the intellectual property embodied in online courses. The American Association of University Professors (AAUP) has taken the position that professors rather than institutions should retain primary property rights for online course materials (Schneider 1999) and has questioned the accreditation of Jones International University (Olsen 1999).

The issues raised by IT in education are still in their infancy and will probably take years to resolve.

## IT, Research, and Knowledge Creation

Information technology is having broad and substantial effects on research and the creation of knowledge. IT facilitates:

- ♦ new ways of communicating and storing scholarly information;
- ♦ new methods of research and new fields of science; and
- ♦ new forms of scientific collaboration.

The effects of IT on research and knowledge creation are important for two reasons. First, they have significant effects on the research community, which in turn affects innovation and education in society. Second, many applications of IT that have been used first in the research community, such as e-mail and the World Wide Web, have later diffused more widely and have had major effects outside of the research community.

## Scholarly Communication

In his 1945 *Atlantic Monthly* article, Vannevar Bush illustrated how helpful it would be to researchers to have access at their desk to the great body of the world's knowledge. In the past few years, that vision has come much closer to reality. The Internet and the World Wide Web, originally developed as tools for scientific communication, have become increasingly powerful. An increasing amount of scholarly information is stored in electronic forms and is available through digital media—primarily the World Wide Web.

Scholars derive many advantages from having scholarly information in digital form. They can find information they want more easily using search tools. They can get the information without leaving their desks, and they do not have to worry about journals being missing from the library. They can get more complete information because electronic publications are not constrained by page limits as printed journals commonly are. Multimedia presentations and software can be combined with text, enriching the information and facilitating further work with it. Additional references, comments from other readers, or communication with the author can be a mouse-click away.

There are also advantages for libraries. Many patrons can access the same electronic information at the same time, possibly without having to visit the library facility; electronic archives do not take up the space held by old journal collections; and libraries can stretch limited financial resources, especially for accessions. All of these factors exert strong pressures for making scholarly information available electronically.

The traditional system of printed academic journals, however, performs functions other than information transmission. Journals organize articles by field and manage peer review processes that help to screen out bad data and research. Scholars achieve recognition through publication in prestigious journals, and universities base hiring and promotion decisions on publication records. Similarly, traditional libraries have played roles in scholarship that go beyond storing books and journals. The library is a place for students and scholars to congregate, and it often has been the intellectual center of a university.

Electronic publications also raise issues about the archiving of information. Rapidly changing IT means that publications stored in one format may not be readily accessible to future users. This problem may become increasingly difficult when electronic “publications” include hyperlinks, multimedia presentations, or software programs.

There are several different ways to put scholarly information online, all of which are expanding. These “media” include individual Web pages, preprint servers, electronic journals, and electronic versions of print journals.

Many scholars put their own work on personal or research-group Web pages. These sites may include “reprints” of published material, preprints, working papers, talks and other unpublished material, bibliographies, data sets, related course material, and other information of use to other scholars. This approach provides an efficient way for scholars to respond to requests for information from colleagues or students.

Another rapidly growing form of electronic publication has been preprint or reprint servers, whereby authors in a specified field post their articles. These servers enable readers to find papers of interest, accelerate dissemination of new knowledge, and provide a focal point for information in a field. The original and most widely copied preprint server is the Los Alamos physics preprint server (<<<http://xxx.lanl.gov/>>>). This site was started in 1991 by Los Alamos physicist Paul Ginsparg as a service to a small subfield of physics; it has grown to cover many fields of physics, astronomy, mathematics, and computation. By mid-1999 it was receiving more than 2,000 new submissions each month and had close to 100,000 connections each day (e.g., for searching, reading, or downloading papers) from approximately 8,000 different hosts. (See figures 9-17 and 9-18.) It has become the main mode of communication in some fields of physics. Fourteen other places around the globe have established mirror sites that copy the information on the Los Alamos server to provide alternative access to the information. One effect of the server is that physicists around the world who do not have access to major research libraries can keep abreast of the latest developments in physics.

The preprint server is a very efficient mode of communication. Odlyzko (1997) estimates that the Los Alamos server costs \$5–\$75 per article (the upper estimate is based on deliberately inflated assumptions about costs), compared to costs of \$2,000–\$4,000 per article for an average scholarly print journal. The server does not provide refereeing of articles, but it does provide a means for scientists to comment on papers that are posted as well as to respond to the comments of others. It also provides a forum for electronic discussions in various fields. The Los Alamos server is frequently regarded as a model. Other preprint servers modeled after the Los Alamos server include the Economics Working Paper Archive hosted by the Economics Department of Washington University (<<<http://econwpa.wustl.edu/wpawelcome.html>>>) and a Chemical Physics Preprint Database operated by the Department of Chemistry at Brown University and the Theoretical Chemistry and Molecular Physics Group at the Los Alamos National Laboratory (<<<http://www.chem.brown.edu/chem-ph.html>>>). As other preprint servers develop, it will become easier to understand how much the Los Alamos success derives from the particular nature of the research and researchers in physics and how much can be generalized.

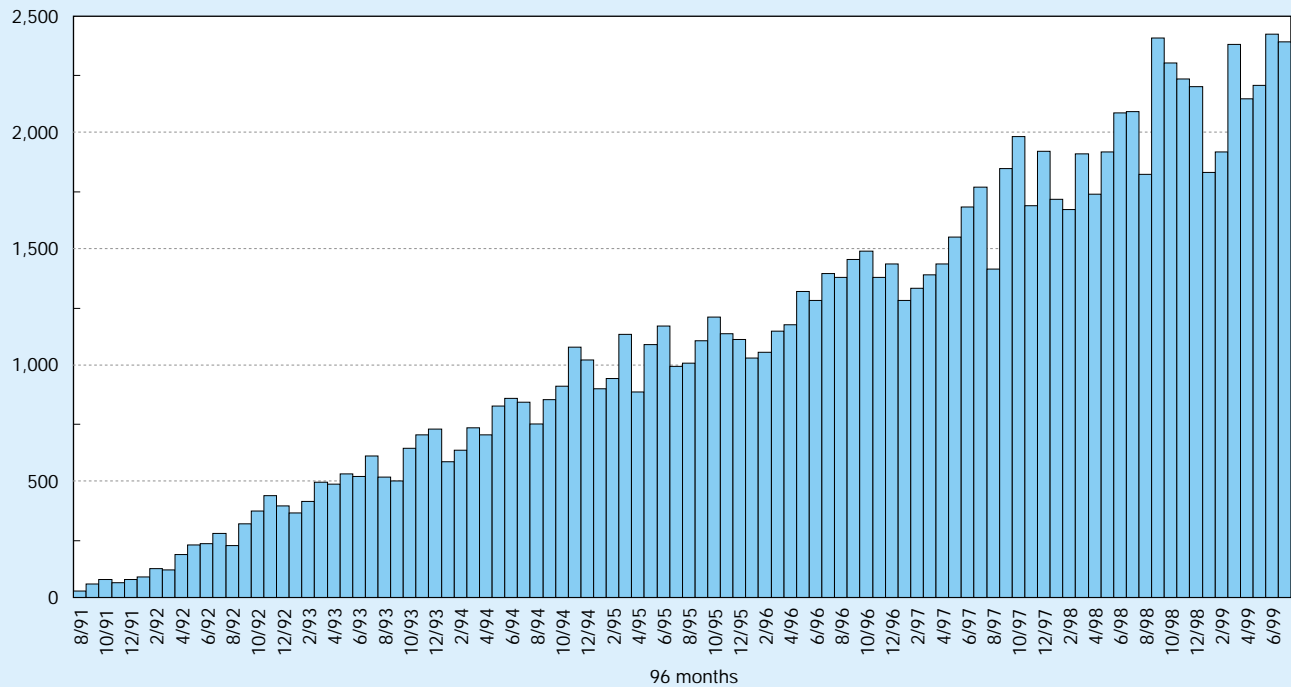
Implementation issues associated with scholarly electronic publishing were underscored by the 1999 proposal by NIH director Harold Varmus for a Web-based repository of biomedical literature to be hosted by NIH originally called E-biomed (Varmus 1999). In the original proposal, this repository was intended to be a preprint server, modeled after the Los Alamos server; that proposal was revised, however, after extensive public comment and discussion in the press. Some people expressed concern that unrefereed medical publications might be a public health risk. Others suggested that NIH, as the funding agency for biomedical research, should not itself publish research results. Much of the criticism came from professional societies and the publishers of academic journals, who regarded E-biomed as a threat to their circulation and revenue. In response to these comments, NIH revised the proposal to create a “reprint” server that would work with existing journals to post the text of those journals after they are published. (NIH also changed the name, first to E-biosci, and then to PubMed Central.) Although this proposal is less threatening to publishers, the benefits to them of participation are not yet clear (Marshall 1999).

The controversy over the Varmus proposal shows that key players include not only researchers and publishers but also the broader public that may access electronic publication. Research posted on the Web that has direct public health or policy implications is likely to receive more scrutiny than research with a primarily scientific audience. As regulatory attention to health information on Web sites illustrates, the quality of some kinds of information may trigger more concern—and intervention—than others.

Electronic journals have also been expanding rapidly. The Association of Research Libraries’ (ARL) 1997 directory of electronic journals, newsletters, and academic discussion lists included 3,400 serial titles—twice as many as in 1996. Of

Figure 9-17.  
Number of new submissions received at Los Alamos preprint server each month since August 1991

Monthly submission rate for archive

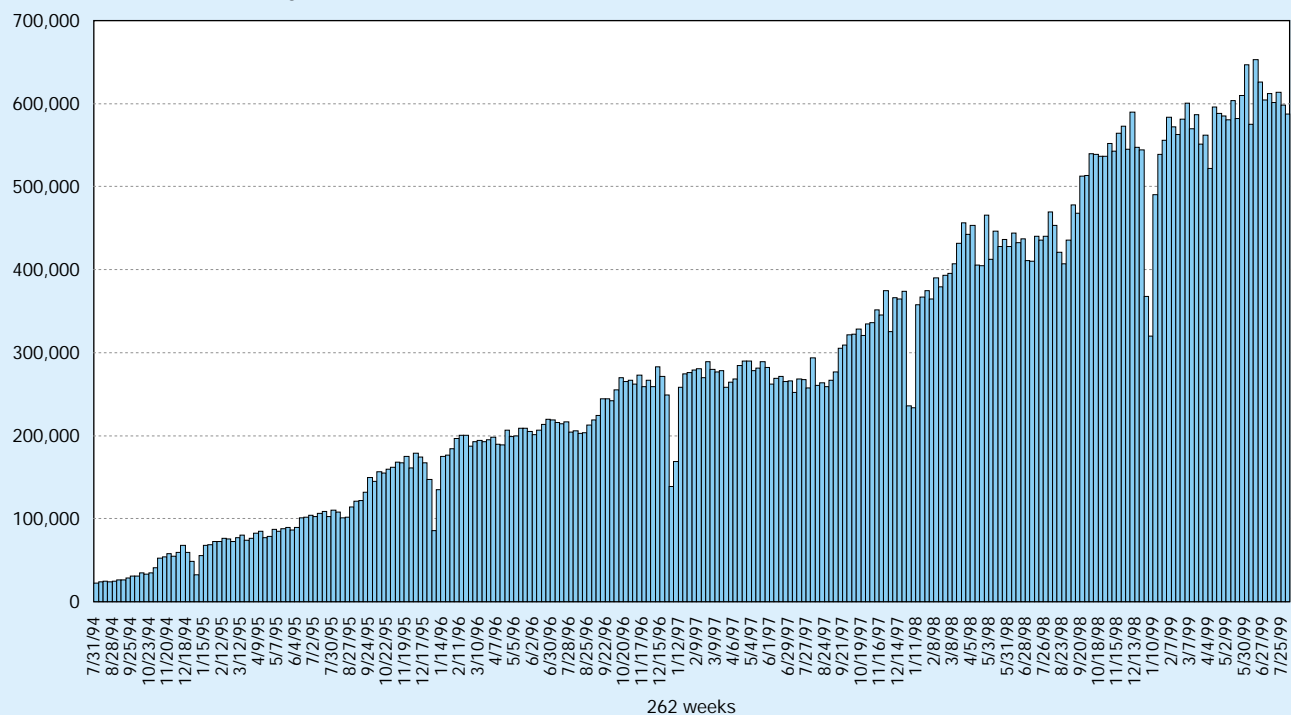


SOURCE: Los Alamos preprint server available from <<<http://xxx.lanl.gov/>>>.

Science & Engineering Indicators – 2000

Figure 9-18.  
Number of connections each week at Los Alamos preprint server: 7/31/94–8/1/99

Number of connections (excluding mirror sites, FTP, e-mail)



SOURCE: Los Alamos preprint server available from <<<http://xxx.lanl.gov/>>>.

Science & Engineering Indicators – 2000

that total, 1,465 titles were categorized as electronic journals; of these, 1,002 were peer-reviewed, and 708 charge in some manner for access. The number of peer-reviewed electronic publications (which includes some publications not classified as journals) has increased rapidly since 1991. (See figure 9-19.) The 1999 ARL directory is expected to list more than 3,000 peer-reviewed titles (Mogge 1999). The increase reflects the fact that traditional print publishers are moving to make their titles available electronically—both as electronic versions of their paper products and as electronic supplements or replacements for the print journal.

Electronic journals can be offered either directly by publishers or through intermediary services that aggregate the titles from many publishers in one service (Machovec 1997). Publishers are currently experimenting with different ways of pricing electronic journals. Some provide separate subscriptions for electronic versions that may be higher or lower cost than the print version. Others provide the electronic version at no charge with a subscription to the print version. Some publishers offer free online access to selected articles from the print version and regard the online version as advertising for the print version (Machovec 1997). Publishers of fee-based electronic journals generally protect their information from unauthorized access by restricting access to certain Internet domains (such as those of universities that have acquired a site license) or through passwords.

Print publishers who move to electronic publishing have found that their costs remain significant (Getz 1997). A large proportion of the cost of most journals covers editing and refereeing of manuscripts and general administration—which, at least initially, remains about the same for electronic journals. In addition, there are costs associated with new information technology and with formatting manuscripts for

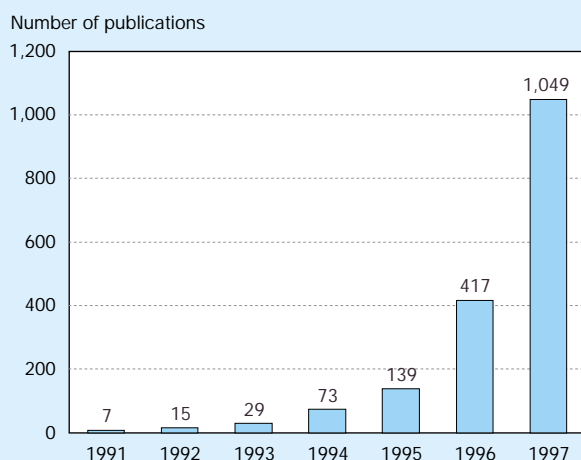
electronic publication. Some of these costs might decline with time, experience, and improved technology.

Electronic publication also can affect the revenue stream of print publishers. If a publisher provides a site license for a university library that enables anyone on campus to read the journal, individual subscriptions may decline. Moreover, electronic journals may be less attractive to advertisers than print versions.

Some electronic-only journals—generally run by an unpaid editor and distributed on the Internet at no cost to the user—are operated at low cost. They can provide a similar filtering function to that of print journals (using, as do other scholarly journals, unpaid reviewers), but they generally have lower administrative and publishing costs. Many free journals are subsidized, directly or indirectly, by another organization; some charge authors fees for articles that are printed to cover their costs. Odlyzko (1997) estimates that these journals can operate at \$250–\$1,000 per article (again, compared to \$2,000–\$4,000 per article for average academic publications).

The system of scholarly communication is changing rapidly, but the direction of that change remains uncertain. Although scholars want to be able to access information in electronic form, and the costs of electronic publishing can be lower, there are some barriers to electronic publishing. Scholars, who do not directly bear the cost of journals, tend to submit their articles to print journals rather than electronic journals because they still regard print journals as more prestigious (Kiernan 1999). (They may also post their articles on the Web for convenience.) Research libraries, which are under pressure to cut journal costs, also must continue to meet the needs of their research communities to provide access to the most important journals (which are mostly still print journals), and libraries have trouble affording print and electronic versions of the same journals. Libraries are seeking new strategies, such as negotiating university-system wide packages for electronic journals to lower costs (Biemiller 1999) or even supporting new, lower cost journals to compete with high-cost journals (ARL 1999).

Figure 9-19.  
Peer-reviewed electronic publications



SOURCE: Mogge, D., *ARL Directory of Electronic Journals, Newsletters and Academic Discussion* (1997): Foreword. Available from <<<http://www.arl.org:591/foreword.html>>>.

Science & Engineering Indicators – 2000

## Digital Libraries

The term “digital library” does not refer to a library in the conventional sense of a central repository of information. Rather, the term encompasses a broad range of methods of storing materials in electronic format and manipulating large collections of those materials effectively. Some digital library projects focus on digitizing perishable or fragile photographs, artwork, documents, recordings, films, and artifacts to preserve their record and allow people to view items that could otherwise not be displayed publicly. Others are digital museums, which allow millions of individuals access to history and culture they would not otherwise have.

One example is JSTOR, an Andrew W. Mellon Foundation-funded project to convert the back issues of paper journals into electronic formats (JSTOR 1999). The goals of this project are to save space in libraries, to improve access to journal content, and to solve preservation problems associ-

ated with storing paper volumes. High-resolution (600 dpi) bit-mapped images of each page are linked to a text file generated with optical character recognition software to enable searching. JSTOR does not publish current issues of the journals, which would put journal publishers' revenue stream at risk; instead, it publishes volumes when they are either three or five years old, depending on the journal. JSTOR now covers more than 117 key journal titles in 15 disciplines. Access to JSTOR is available through institutions such as university libraries that have site licenses.

The Federal Government's multi-agency Digital Library Initiative (<<<http://www.dli2.nsf.gov/>>>) is supporting projects at many universities around the country. These projects are designed to improve methods of collecting, storing, and organizing information in digital forms and to make information available for searching, retrieval, and processing via communication networks. These projects cover a broad range of topics in the sciences, social sciences, arts, and humanities. They cover information creation, access and use, and archiving and preservation for information as diverse as maps, videos, scientific simulations, and medical records. That diversity enriches the IT through these projects and the clientele for electronic information. It also differentiates digital library projects from preprint servers. The sidebar "Growth of the World Wide Web" provides additional information on libraries and the Web.

## Effects of IT on Research

IT has had a major effect on research. It has facilitated new methods of research and development, new forms of research collaboration, and new fields of science. Computers have affected research from their beginnings, and scientific users historically have had the most advanced computing capability. Today, advances in the underlying technology make relatively advanced capabilities available more broadly, fueling the diffusion of IT from its historical stronghold in the physical sciences across the research community through other natural sciences, engineering, social sciences, and the humanities.

### New Research Methods

High-end computing and software have had a fundamental impact on research in many areas of science and technology. Some areas of research—such as high-energy physics, fluid dynamics, aeronautical engineering, and atmospheric sciences—have long relied on high-end computing. The ability to collect, manipulate, and share massive amounts of data has long been essential in areas such as astronomy and geosphere and biosphere studies (Committee on Issues in the Transborder Flow of Scientific Data 1997). As information technologies have become increasingly powerful, they have facilitated continued advances in these areas of science and become increasingly vital to sciences such as biology that historically used IT less extensively.

Shared databases have become important resources in many fields of science and social sciences. Examples include

Census Bureau databases, data from large scientific instruments such as the Hubble Space Telescope, genetic and protein databases (e.g., GenBank), and the NIH-funded human brain project, as well as many smaller and more specialized databases. These databases allow researchers working on different pieces of large problems to contribute to and benefit from the work of other researchers and shared resources.

Modeling and simulation have become powerful complements to theory and experimentation in advancing knowledge in many areas of science. Simulations allow researchers to run virtual experiments that, for either physical or practical reasons, they cannot run in reality. As computer power grows, simulations can be made more complex, and new classes of problems can be realistically simulated. Simulation is contributing to major advances in weather and climate prediction, computational biology, plasma science, high-energy physics, cosmology, materials research, and combustion, among other areas. Industry also uses simulations extensively to test the crashworthiness of cars and the flight performance of aircraft (DOE/NSF 1998) and to develop new financial instruments (e.g., derivatives).

The performance of computers continues to improve at a rapid rate. The Department of Energy's Accelerated Strategic Computing Initiative program, which uses simulation to replace nuclear tests, deployed the first trillion-operations-per-second (teraops) computer in December 1996 and is planning to operate a 100 teraops computer by 2004 (National Science and Technology Council 1999). Researchers funded by DARPA, NASA, and the National Security Agency (NSA) are evaluating the feasibility of constructing a computing system capable of a sustained rate of  $10^{15}$  floating point operations per second (1 petaflop).

### IT in Biology

IT is becoming increasingly important in biology. Genomics research, including efforts to completely map the human genome (which consists of 3 billion nucleotide base pairs) by 2005, depends on robots to process samples and computers to manage, store, compare, and retrieve the data (Varmus 1998). The databases that contain gene and protein sequence information have been growing at an enormous rate. GenBank, NIH's annotated collection of all publicly available DNA sequences, has been growing at an exponential rate: The number of nucleotide base pairs in its database has been doubling approximately every 14 months. As of August 1999, GenBank contained approximately 3.4 billion base pairs, from 4.6 million sequence records. These base pairs were from 50,000 species; *Homo sapiens* accounted for 1.8 billion of the base pairs. (See figure 9-21.)

GenBank is part of a global collaboration; it exchanges data daily with European and Japanese gene banks. In addition to the publicly available sequences in GenBank, private companies are rapidly developing proprietary genetic sequences.

To make use of data from the human genome project, new computational tools are needed to determine the three-dimensional atomic structure and dynamic behavior of gene products, as well as to dissect the roles of individual genes

## Growth of the World Wide Web

One indicator of the growth of digital information is the growth of the World Wide Web. The volume of information on the Web has grown enormously. (See figure 9-20.) Although scholarly information is only a small part of the Web, the amount of useful scholarly information is still large.

Lesk (1997a) notes that a book such as *Moby Dick* is approximately 1 megabyte in plain-text ASCII form, so 1 terabyte is the equivalent of 1 million substantial books. By this measure, the amount of text on the Web as of February 1999 was equivalent to 6 million books.

Lawrence and Giles (1999) estimate that there were 800 million pages on the publicly indexable Web as of February 1999—corresponding to 15 terabytes in HTML or 6 terabytes in text.\* They also estimated that 3 terabytes of image data were available online. They found that about 6 percent of Web servers have scientific or educational content—defined as university, college, or research lab servers.

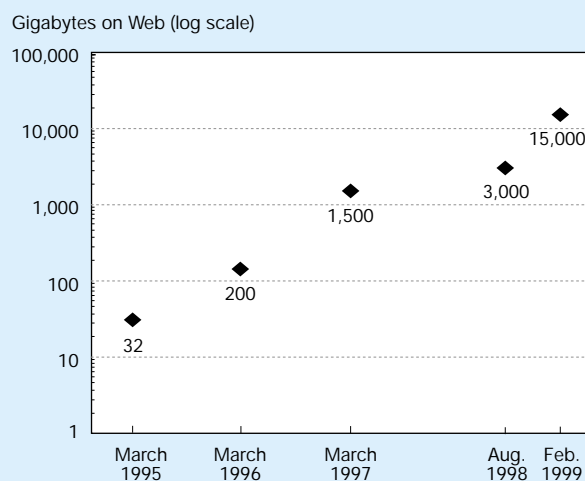
In addition to the World Wide Web, other online information providers such as Dialog and Lexis-Nexis make large amounts of information available. Dialog has approximately 9.2 terabytes and Lexis-Nexis has approximately 5.9 terabytes (Lesk 1997a). Many universities now have access to Lexis-Nexis (Young 1998).

By comparison, the largest library in the world, the Library of Congress, has 17 million books—equivalent to 17 terabytes of text. The Library of Congress also has 2 million recordings, 12 million photographs, 4 million maps, 500,000 films, and 50 million manuscripts. In all, it has 115 million items (Library of Congress 1999). Because these other types of collection would be very large in digital form, the collections in the Library of Congress might total 3,000 terabytes (Lesk 1997a).

Thus, the amount of information in network-accessible digital form is already very large and is approaching the volume of text in the largest libraries. It already exceeds the volume of text in libraries that are readily accessible to most people. It does not yet, however, match the total holdings of the largest libraries in sheer volume. On the other hand, the range of information available online is broader than that in most libraries, albeit in ways that do not necessarily make it more useful—as typical results of Web searches illustrate today. The amount of information available online is growing quickly and will likely grow even faster as more people obtain higher-bandwidth Internet connections and can more readily use the Internet for music, video, and multimedia information that they generate as well as consume.

Of course, there are great qualitative differences between material in libraries and material on the Web. Most material in libraries has been judged by editors and librarians to have some lasting value—it has been selected. Much of the material on the Web has not gone through such filters and has been generated for a wider variety of purposes (e.g., public relations or commercial information). In addition, for most of the material on the Web, there is no guarantee that the information will be accessible in the future. On the other hand, the Web is useful as a source for materials such as preprints and technical reports that may be difficult to find in libraries.

Figure 9-20.  
Growth in number of gigabytes on the Web



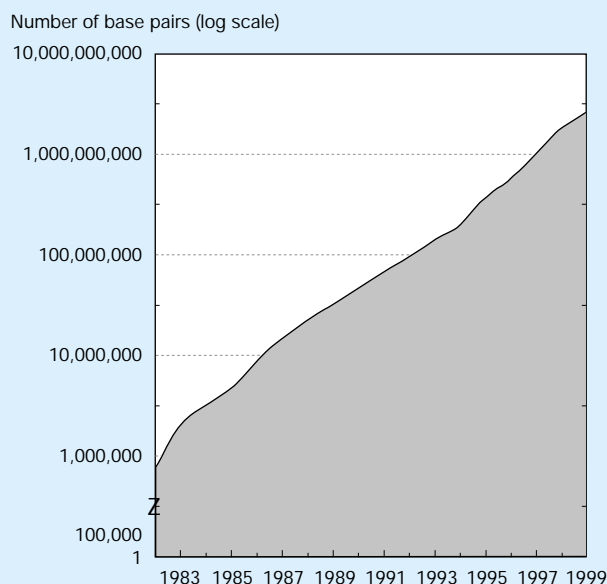
NOTE: The larger jump from 1998 to 1999 may be because Alexa counted actual pages it found and retrieved, whereas Lawrence and Giles used a sampling technique.

SOURCES: 1996, 1997, and 1998 data from Alexa <<www.alexa.com/company/inthenews/webfacts.html>>; 1999 data from Lawrence, S., and L. Giles, "Accessibility of Information on the Web," *Nature* 400 (July 8): 107–109.

*Science & Engineering Indicators – 2000*

\*Lawrence and Giles tested 3.6 million random Internet Protocol (IP) addresses to see if there was a server at that address. They found one server for every 269 requests. Because there are 4.3 billion possible IP addresses, this result led to an estimate of 16 million Web servers. After eliminating servers that were not publicly indexable (such as those behind firewalls or those with no content), they estimated the publicly indexable Web to comprise 2.8 million servers. Lawrence and Giles sampled 2,500 of these servers at random and found the average number of pages per server to be 289, leading to an estimate of 800 million Web pages. These pages averaged 18.7 kilobytes (7.3 kilobytes of text after HTML tags were removed). Lawrence and Giles also found 62.8 images per server, with a mean size of 15.2 kilobytes. Using a similar sampling method, the Online Computer Library Center (OCLC 1999) estimated that there were 288 million ( $\pm 35$  percent) unique, publicly accessible Web pages in June 1999.

Figure 9-21.  
Growth of Genbank



SOURCE: Genetic Sequence Data Bank, NCBI-GenBank Flat File Release 113.0 (15 August 1999); Distribution Release Notes. Available from <<ftp://ncbi.nlm.nih.gov/genbank/gbrel.txt>>.

Science & Engineering Indicators – 2000

and the integrated function of thousands of genes. To model the folding of a protein—a capability that would dramatically aid the design of new drug therapies—takes the equivalent of months of Cray T3E computer time (DOE/NSF 1998). Researchers are also using pattern recognition and data mining software to help decipher the genetic information (Regalado 1999).

The importance of informatics for biology and medicine is difficult to overemphasize. Many scientists expect it to revolutionize biology in the coming decades, as scientists decode genetic information and figure out how it relates to the function of organisms. As NIH director Varmus (1999) stated, “All of biology is undergoing fundamental change as a result of new methods that permit the isolation, amplification, and detailed analysis of genes.” Genomic information will be used to assess predisposition to disease, predict responses to environmental agents and drugs, design new medicines and vaccines, and detect infectious agents. New areas of biology—such as molecular epidemiology, functional genomics, and pharmacogenetics—rely on DNA data and benefit more generally from new, information-intensive approaches to research.

### Research Collaboration

IT facilitates enhanced collaboration among scientists and engineers. E-mail, the World Wide Web, and digital libraries allow information to be accessed from anywhere and let geographically separated scientists (even if they

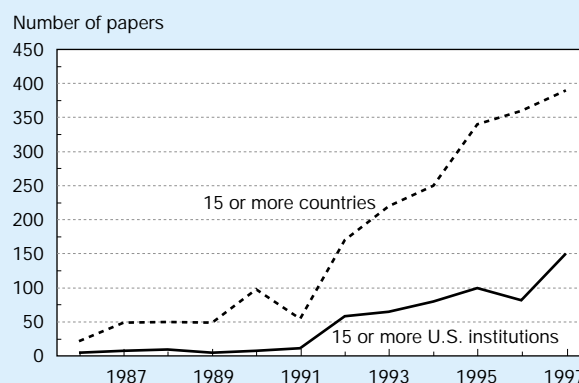
are only a building away) work together better. Some companies with laboratories around the world pass off problems from one lab to another so researchers can work on the problems 24 hours a day.

Scientific collaboration—as measured by the increase in the percentage of papers with multiple authors—has been increasing steadily for decades. Much of this collaboration is probably the result of better telephone service and air travel, as well as the availability of fax machines and e-mail. Large-scale scientific collaborations may be especially enabled by new information technology. There has been a rapid increase in the number of papers with authors from many institutions that coincides with the rapid expansion of the Internet. (See figure 9-22.)

More advanced technologies to aid R&D collaboration are coming into use and are likely to migrate to broader usage in the next few years. (See sidebar, “Collaboratories.”)

How the application of IT will affect the science and engineering enterprise in the long run is not clear. Although the potential for change is obvious, we do not know how much and what kind of change will endure. The availability of information from anywhere may reduce the need for researchers to be close to major research libraries. The ability to operate major scientific instruments over the Web may reduce the need for scientists to be located at major laboratories. If virtual laboratories can function effectively, there may be less need to assemble large multidisciplinary teams of scientists and engineers at a laboratory to work on complex problems at a common location. Most scientists, however, may still want extensive face-to-face interaction with their colleagues, and they may want hands-on participation in experiments.

Figure 9-22.  
Number of papers with authors from 15 or more countries or 15 or more U.S. institutions: 1986–97



SOURCE: CHI Research, Inc.

Science & Engineering Indicators – 2000



## Collaboratories

In 1989, William Wulf (now president of the National Academy of Engineering but then at the National Science Foundation) coined the term “collaboratory” to describe the concept of using information technologies to make geographically separate research units function as a single laboratory. Wulf defined a “collaboratory” as a “...‘center without walls’ in which the nation’s researchers can perform their research without regard to geographical location—interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information in digital libraries” (CSTB 1993).

In subsequent years, a number of programs began to develop tools for collaboratories and fund pilot projects. Among the earliest projects were:

- ◆ The NSF-sponsored Upper Atmosphere Research Collaboratory (UARC)—now the Space Physics and Aeronomy Research Collaboratory (SPARC)—which allows space physics researchers around the world to control and gather data from more than a dozen instruments located around and above the globe. SPARC is based at University of Michigan (<<<http://www.crew.umich.edu/UARC/>>>); it has collaborators from many institutions.
- ◆ The DOE-sponsored Materials MicroCharacterization Collaboratory (<<<http://tpm.amc.anl.gov/MMC>>>), which conducts research on the microstructure of advanced materials. This effort involves three DOE national laboratories, the National Institute for Standards and Technology (NIST), the University of Illinois, and several scientific instrument companies.
- ◆ The DOE-sponsored Diesel Combustion Collaboratory (<<<http://www.collab.ca.sandia.gov/Diesel/ui/>>>), which focuses on diesel engine emissions control and involves three DOE national laboratories, the University of Wisconsin, and several diesel engine manufacturers.

These collaboratories use a similar set of technologies for collaboration, including:

- ◆ Internet-based desktop video conferencing;
- ◆ Shared access to databases and computer simulation;

- ◆ Shared virtual workspaces, such as “white boards” on which researchers can sketch out ideas; and
- ◆ Shared electronic laboratory notebooks to capture the details of experiments.

One of the most important aspects of collaboratories is the ability to share scientific instruments over the Internet. This sharing may involve many users from different sites using a single major scientific instrument, such as a synchrotron at a national laboratory, or it may involve using a network of instruments, such as environmental sensors in geographically separate parts of the globe.

Many of the tools developed in these and other pilot projects are now being used in other research collaborations.\*

Among the benefits of collaboratories (Ross-Flanigan 1998) are that:

- ◆ Scientists can avoid going to scientific instruments in remote locations.
- ◆ Many more universities, scientists, and students can participate in or observe experiments.
- ◆ By connecting computation to experiments, scientists can better and more quickly integrate experiments and theory. Theorists and experimentalists can work together in real time, greatly reducing the time required to analyze experiments.
- ◆ Scientists can put together quick video conferences to discuss the data.
- ◆ Students can participate in experimentation much earlier in their careers than before.

On the other hand, virtual communication has been found to be less successful than face-to-face communication in building trust between researchers. In addition, as a result of greater outside participation in the research, good researchers have more distractions. The early collaboratories also found that Internet congestion, the lack of reliability of some of the tools, and software changes slowed research.

\*See, for example, <<<http://www.si.umich.edu/research/projects.htm#collabor>>>; <<<http://www.mcs.anl.gov/DOE2000/pilot.html>>>; <<<http://doe2k.lbl.gov/doe2k/index.html>>>.

## IT and the Citizen

### IT in the Home

The breadth of information technologies in the home is considerable, ranging from televisions and telephones to smart house technology, microprocessors in coffee pots, personal computers (PCs), and the Internet.<sup>8</sup> The trends and develop-

ments presented here focus only on home computers and Internet linkages,<sup>9</sup> not on the full spectrum of home informatics or ways in which people can access the Internet outside of the office (such as in libraries, kiosks, or Internet cafes). In addition, the analysis concentrates on social impacts that occur within the home itself, such as changes in individuals, in family dynamics, or in household operations.

<sup>8</sup>For a more extensive discussion of the diffusion and effects of information technologies in the home, see National Science Foundation, *The Applications and Implications of Information Technologies for the Home* (1999) (available at <<[http://srweb.nsf.gov/it\\_site/index.htm](http://srweb.nsf.gov/it_site/index.htm)>>); National Technical Information Administration, *Falling Through the Net: Defining the Digital Divide* (1999) (available at <<<http://www.ntia.doc.gov/ntiahome/digitaldivide/>>>).

<sup>9</sup>Note that there is increasing diversity in technical access to the Internet—for example, through television (web TV<sup>TM</sup>) and telephones. Such alternative mechanisms are not explicitly addressed in this study; most research reviewed here assumes that Internet access is achieved through a personal computer.